

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 03:18:06 ; Search time 9.6 Seconds
(without alignments)
1581.285 Million cell updates/sec

Title: US-09-807-933B-3

Sequence: 1 MKRTITSSALLALALGTEM.....TYKEVCPKRTAKTGSRK.366
2020

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683.5	33.8	376	1 GUNK_FUSOX	P45699 fusarium ox
2	674	33.4	213	1 GUN5_HUMIN	P43316 humicola in
3	552	27.3	511	1 GUNB_PSEPL	P18126 pseudomona
4	511.5	25.3	393	1 GUN1_USPTA	P54424 useliago ma
5	238.5	11.8	210	1 PSBP_PORPU	P50272 porphyra pu
6	227.5	11.3	471	1 GUX2_TRIKE	P07987 trichoderma
7	177.5	8.8	438	1 GUX3_AGABI	P49075 agaricus bi
8	167.5	8.3	418	1 GUN2_TRIRE	P07982 trichoderma
9	158.5	7.8	540	1 GUX1_ASAPC	O59843 aspergillus
10	154	7.6	242	1 GUN5_TRIRE	P43317 trichoderma
11	152	7.5	462	1 GUNB_FUSOX	P46236 fusarium ox
12	147.5	7.3	1218	1 JAG1_HUMAN	P78504 homo sapien
13	147.5	7.3	1218	1 JAG1_MOUSE	O99430 mus musculu
14	145	7.2	485	1 Y136_TREPA	O83172 trepomena p
15	145	7.2	1242	1 JAG1_BRAAE	O90957 brachydanio
16	144.5	7.2	1213	1 JAG3_BRAAE	O90954 brachydanio
17	144	7.1	442	1 CYS4_DICDI	P54639 dictyosteli
18	143	7.1	385	1 GUNF_FUSOX	P46239 fusarium ox
19	141.5	7.0	1219	1 JAG1_RAT	O63722 rattus norv
20	140.5	7.0	506	1 GUX2_AGABI	O92400 agaricus bi
21	140.5	7.0	713	1 TSA4_GIALA	P12849 giardia lam
22	139.5	6.9	5376	1 ZAN_MOUSE	O88799 mus musculu
23	139	6.9	2704	1 G16B_PARRR	P17053 paramecium
24	135.5	6.7	388	1 GUN3_HUMIN	Q12624 humicola in
25	135	6.7	513	1 GUX1_TRIPI	P19355 trichoderma
26	134	6.6	513	1 GUX1_TRIPI	P00725 trichoderma
27	134	6.6	1202	1 JAG2_RAT	P97607 rattus norv
28	134	6.6	2421	1 NTC2_RAT	O99430 rattus norv
29	134	6.6	2703	1 NOTC_DROME	P07207 drosophila
30	133.5	6.6	1700	1 BAR3_CHITE	O93376 chironomus
31	133	6.6	2715	1 G156_PARRR	P13837 paramecium
32	132	6.5	2471	1 NTC2_HUMAN	Q04721 homo sapien
33	129.5	6.4	355	1 YW79_YEAST	Q04018 saccharomyc

34	129.5	6.4	687	1 VS41_GIALA	P92127 giardia lam
35	129	6.4	344	1 GUN4_TRIRE	O14405 trichoderma
36	129	6.4	1246	1 YWV2_CABEL	P34504 caenorhabdl
37	129	6.4	1408	1 SER1_DROME	P18168 drosophila
38	129	6.4	2531	1 NTC1_MOUSE	O01705 mus musculu
39	128.5	6.4	537	1 SP70_DICDI	P15269 dictyosteli
40	128	6.3	667	1 TS11_GIALA	O03185 giardia lam
41	128	6.3	833	1 DL_DROME	P10041 drosophila
42	128	6.3	3712	1 LMA_DROME	O00174 drosophila
43	127.5	6.3	227	1 AGI_ORYSA	P11219 oryza sativ
44	127.5	6.3	316	1 LORI_HUMAN	P23490 homo sapien
45	127.5	6.3	1238	1 JAG2_HUMAN	O92119 homo sapien

ALIGNMENTS

RESULT 1	GUNK_FUSOX	STANDARD:	PRT:	376 AA.
ID	GUNK_FUSOX			
AC	P45699;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).			
OS	Fusarium oxysporum.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreales; mitosporic Hypocreales; Fusarium.			
OX	NCBI_TaxID=5507;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95047531; PubMed=7959045;			
RA	Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,			
RA	Hagen F.S., Upshall A., McKnight G.V., O'Hara P.J.;			
RT	"The use of conserved cellulase family-specific sequences to clone			
RT	cellulase homologue cDNAs from Fusarium oxysporum.";			
RL	Gene 150:163-167(1994).			
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).			
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL			
CC	HYDROLASES).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.1ib-sib.ch/announce/			
CC	or send an email to license@lib-sib.ch).			
CC	-----			
DR	EMBL; L29381; AAA65589.1; -.			
DR	HSSP; P43316; 2ENG.			
DR	InterPro: IPR000254; CBD fungal.			
DR	InterPro: IPR000334; GH_45.			
DR	Pfam; PF00734; CBM_1; 1_45.			
DR	Pfam; PF02015; Glyco_hydro_45; 1.			
DR	SMART; SM00236; FCBP; 1.			
DR	PROSITE; PS00562; CBD_FUNGAL; 1.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
FT	SIGNAL	1	18	POTENTIAL
FT	CHAIN	19	376	PUTATIVE ENDOGLUCANASE TYPE K.
FT	DOMAIN	19	308	CATALYTIC.
FT	DOMAIN	309	338	LINKER.
FT	DOMAIN	339	376	CELLULOSE-BINDING.
FT	ACT_SITE	29	29	NUCLEOPHILE (BY SIMILARITY).
FT	ACT_SITE	140	140	PROTON DONOR (BY SIMILARITY).
SO	SEQUENCE	376 AA;	39235 MW;	B430A5F962B9F882 CRC64;

Query Match 33.8%; Score 683.5; DB 1; Length 376;
Best Local Similarity 57.1%; Pred. No. 1.2e-38;

Matches 120; Conservative 34; Mismatches 49; Indels 7; Gaps 4;

QY 157 ASGNVTRRYMDCCASGSPGKANVSPYKSCNKDGTALSDSNVSGCNGNS-YMGN 215
 DB 18 ASGSHSTRYMDCCASGSPGKANVSPYKSCNKDGTALSDSNVSGCNGNS-YMGN 215
 QY 216 DNQPAVNDNLAYGAAALISGGESRWCCSCFELTFTSTSVAGKMMIQTNTGDLGS 275
 DB 77 NYSFPAVNDNLAYGAAALISGGESRWCCSCFELTFTSTSVAGKMMIQTNTGDLGS 275
 QY 276 SGAHFDLQMPGGGAGIFNGSKWGA-PNDGMSRRYGGISSASDSSLPALQAGCKMRF 335
 DB 137 N---HFDMLEGGGAGIFNGSKWGA-PNDGMSRRYGGISSASDSSLPALQAGCKMRF 335
 QY 336 NMFKNADNPMTYKEVTCPEKITAKTGCGR 365
 DB 192 DMFENADNPMTYKEVTCPEKITAKTGCGR 221

RESULT 2

GUNS_HUMIN STANDARD; PRT; 213 AA.

AC P43316;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
 DB (Cellulase V) (EG V).
 OS Humicola insolens.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 NX NCBI_Taxid=34413;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Rasmussen G., Mikkelsen J.-M., Schuelein M., Packar S.A., Hagen F.,
 RA Hjort C.M., Hestrup S.;
 RL "A cellulase preparation comprising an endoglucanase enzyme."; Patent number WO9117243, 14-NOV-1991.
 RN (2)
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=93390621; PubMed=8377830;
 RA Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,
 RA Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schuelein M.;
 RL "Structure and function of endoglucanase V."; Nature 363:362-364 (1993).
 RN (3)
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=96101453; PubMed=8519779;
 RA Davies G.J., Tolley S.P., Hentisat B., Hjort C., Schuelein M.;
 RL "Structures of oligosaccharide-bound forms of the endoglucanase V from Humicola insolens at 1.9-A resolution."; Biochemistry 34:16210-16220 (1995).
 RN (4)
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RA Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,
 RA Wilson K.S., Rasmussen G., Schuelein M.;
 RL "Structure determination and refinement of the Humicola insolens endoglucanase V at 1.5-A resolution."; Acta Crystallogr. D 52:7-17 (1996).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL HYDROLASES).
 CC PDB; 2ENG; 08-DEC-96.
 DR PDB; 3ENG; 16-JUN-97.
 DR PDB; 4ENG; 16-JUN-97.
 DR InterPro: IPR000334; GH 45.
 DR Pfam; PF02015; Glyco_hydro_45; 1.
 DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
 DR Cellulose degradation; Hydrolyase; Glycosidase; 3D-structure.
 FT ACT SITE 10 10 NUCLEOPHILE.
 FT ACT SITE 121 121 PROTON DONOR.
 SO SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64;

Query Match 33.4%; Score 674; DB 1; Length 213;
 Best Local Similarity 58.2%; Pred. No. 3e-38;
 Matches 121; Conservative 32; Mismatches 45; Indels 10; Gaps 5;

QY 160 NGVTRRYMDCCASGSPGKANVSPYKSCNKDGTALSDSNVSGCNGN-IGNSYMCNDQ 218
 DB 2 DGRSTRYMDCCASGSPGKANVSPYKSCNKDGTALSDSNVSGCNGN-IGNSYMCNDQ 218
 QY 219 FPAVNDNLAYGAAALISGGESRWCCSCFELTFTSTSVAGKMMIQTNTGDLGSSTG 278
 DB 61 FPAVNDNLAYGAAALISGGESRWCCSCFELTFTSTSVAGKMMIQTNTGDLGSSTG 278
 QY 279 AHFDLQMPGGGAGIFNGSKWGA-PNDGMSRRYGGISSASDSSLPALQAGCKMRF 337
 DB 119 -HFDMLEGGGAGIFNGSKWGA-PNDGMSRRYGGISSASDSSLPALQAGCKMRF 337
 QY 338 NMFKNADNPMTYKEVTCPEKITAKTGCGR 365
 DB 174 FMKNADNPMTYKEVTCPEKITAKTGCGR 201

RESULT 3

GUNB_PSEFL STANDARD; PRT; 511 AA.

AC P18126;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DB (Cellulase) (EGB).
 GN CELB.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NX NCBI_Taxid=294;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.
 RC STRAIN=Sp. Cellulosa;
 RX MEDLINE=90355836; PubMed=2117693;
 RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;
 RL "The N-terminal region of an endoglucanase from Pseudomonas fluorescens subspecies cellulosa constitutes a cellulose-binding domain that is distinct from the catalytic centre."; Mol. Microbiol. 4:759-767 (1990).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRYSTALLINE CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; X52615; CAA36844.1; -.
 DR PIR; S10527; S10527.
 DR HSP; P43316; 2ENG.
 DR InterPro; IPR001919; Bac celose-bind.
 DR InterPro; IPR002883; CBD 5.
 DR InterPro; IPR000334; GH 45.
 DR Pfam; PF00553; CBM 2; 1.
 DR Pfam; PF02013; CBM_10; 1.
 DR Pfam; PF02015; Glyco_hydro_45; 1.

DR PROSITE; PS00561; CBD BACTERIAL; 1.
 DR PROSITE; PS01140; GLYCOSYL_HYDROL_P45; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.
 FT SIGNAL 1 29
 FT CHAIN 30 511
 FT DOMAIN 30 511
 FT DOMAIN 132 131
 FT DOMAIN 132 173
 FT DOMAIN 223 259
 FT DISULFID 32 127
 FT ACT_SITE 276 276
 FT ACT_SITE 393 393
 SQ SEQUENCE 511 AA; 52078 MM; 3C3119D98291D8E CRC64;
 Query Match 27.3%; Score 552; DB 1; Length 511;
 Best Local Similarity 34.0%; Pred. No. 8-2e-30;
 Matches 139; Conservative 56; Mismatches 136; Indels 78; Gaps 16;
 QY 17 GTENASAKSCKLKYGQ-CGGKDMNGPT--CCESGSTCKVNDYYSQCLAPESN--GNKSS 71
 DB 110 GNNSSRAQVAVYGAICGGGSSAPSSVSSSSSVSTPSSSSSVSSVPTSSS 169
 QY 72 ECKLYGCGGCKDMNG--PTC-----CESGSTCKVNDYYSQCLAPESNG 114
 DB 170 SSSSVLTGAQACNMYGTLTPLCNNTSNGWGYEDGRSCVARTTCAQAPAYGIVSTSSSTP 229
 QY 115 NKTSESAHKTTTTTAPAKEITTTAKASNSNSGKYSIVSGSAGNGVTYRPMCKKASC 174
 DB 230 LSSSSSSSSSVSSSSSSSST--SSSASSSVSSVPPIDGC--NGATRYPMCKCKPHC 282
 QY 175 SWPKANYS---SPVSKCNKDGVTALSDSNVQSGCNGNSYKCNNDQPMVAVNDILAVFA 231
 DB 283 GW--SANNVPLVSPLOGCSANN--IRLSVSVSGSGDGGGGMCKDIPFAISPTLAYTA 339
 QY 232 AAATSGGSESRWCCSCFELTFT-----STSVAGKKMVTQVNTGSDLSSTGAH 280
 DB 340 ATS--SGD---VCRCYQLQFTGSSYNAPDPSGSHALAKMTMIVQATNIGYDV--SGGQ 391
 QY 281 FDLMPGGGVIPIFGCCSKOMGAPNDGSRGGYSSA-----SDCS 321
 DB 392 FDLVPGGAGVAFNACSNQMGVSNALQAYGFLAACKQQLYNASLSQYKSCVLANCD 451
 QY 322 SL-----PSALQAGCKMRFNFKNADNPMYKVEVTCPEKITPAKTCGR 365
 DB 452 SVFSGRLTQLQGGCTWFAWFEADNPSLKKEVPCALITTSQNMNR 500
 RESULT 4
 GUNI_USTMA STANDARD; PRT; 393 AA.
 ID GUNI_USTMA
 AC P54424;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-Beta-glucanase 1)
 DE (Cellulase 1) (EG 1).
 GN EGL1
 OS Ustilago maydis (Smut fungus).
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
 OX NCBI_TaxID=5270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FED11;
 RX MEDLINE=96145728; PubMed=8590631;
 RA Schaevecker F., Wanner G., Kahmann R.;
 RT "Filament-specific expression of a cellulase gene in the dimorphic
 RT fungus Ustilago maydis";
 RT Biol. Chem. Hoppe-Seyler 376:617-625(1995).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: HYPHAL TIP.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.

CC -1- PTM: MAY ALSO BE O-GLYCOSYLATED.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
 CC HYDROLASES).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>
 CC or send an email to license@ebi-sib.ch).
 CC
 CC EMBL; S81598; AAB36147.1; -
 DR HSSP; P43316; ZENG.
 DR InterPro; IPR000334; GH_45.
 DR Pfam; PF02015; Glyco_hydro_45_1.
 DR PROSITE; PS01140; GLYCOSYL_HYDROL_P45; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.
 FT SIGNAL 1 26
 FT CHAIN 27 393
 FT ACT_SITE 34 34
 FT ACT_SITE 152 152
 FT DOMAIN 270 385
 FT CARBOHYD 343 343
 SQ SEQUENCE 393 AA; 39594 MM; 65C753C610CDAD3 CRC64;
 Query Match 25.3%; Score 511.5; DB 1; Length 393;
 Best Local Similarity 47.0%; Pred. No. 3.1e-27;
 Matches 103; Conservative 30; Mismatches 65; Indels 21; Gaps 9;
 QY 161 GVTTRYNDCCASGMPGKANVSSPVKCNKDGVTAL---SDSNVQSGCNGNSYKCNNDN 217
 DB 27 GMATRYNDCCASGMPGKANVAVPACRAGDVTLLDSDKDPGSGCNGNKNFMKSCM 86
 QY 218 QPW--AVNDILAVFAAAATSGGSESRWCCSCFELTFT---TSTSVAGKKMVTQVNTG 271
 DB 87 QPFDDETPTLAFGF--CAFTTQGESDPTDCAFYAEFEHDAQGMKRNKTLIPVTVNGG 144
 QY 272 DLGSGTAHPLQMPGGGVIPIFGCCSKOMGAPNDGSRGGYSSA--SDCSLSALQAG 330
 DB 145 DVQSQ---NEDFQIPGGGLAFPGKCPAQMGVSNALQAYGFLAACKQQLYNASLSQYKSCVLANCD 201
 QY 331 CKMRFN--WFKNADNPM--TYKEVTCPEKITPAKTCGR 366
 DB 202 CKMRFSEW---GDNPVLKGSPPKRYCKSLIDRSGCGRK 237
 RESULT 5
 PSBP_PORPU STANDARD; PRT; 210 AA.
 ID PSBP_PORPU
 AC P50272;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative polysaccharide binding protein precursor.
 DE Porphyra purpurea.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
 OX NCBI_TaxID=2787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Avonport;
 RA Liu Q., der Meer J.P., Reith M.E.;
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 4 FUNGAL-TYPE CELLULOSE-BINDING DOMAINS
 CC (CBD).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>
 CC or send an email to license@ebi-sib.ch).

```

CC EMBL; U08843; AAA61792.1; -
DR HSSP; P00725; 2CBH.
DR InterPro; IPR000254; CBD_fungal.
DR Pfam; PF00734; CBM_1; 4.
DR SMART; SM00236; FCBP; 4.
DR PROSITE; PS00562; CBD_FUNGAL; 4.
KM Repeat; Signal.
FT SIGNAL 1 22
FT CHAIN 23 210 PUTATIVE POLYSACCHARIDE BINDING PROTEIN.
FT DOMAIN 23 62 POLYSACCHARIDE BINDING DOMAIN I.
FT DOMAIN 63 105 POLYSACCHARIDE BINDING DOMAIN II.
FT DOMAIN 125 165 POLYSACCHARIDE BINDING DOMAIN III.
FT DOMAIN 166 210 POLYSACCHARIDE BINDING DOMAIN IV.
SQ SEQUENCE 210 AA; 21967 MW; 837822D8B6863C8 CRC64;

Query Match 11.8%; Score 238.5; DB 1; Length 210;
Best Local Similarity 28.5%; Pred. No. 2e-09;
Matches 73; Conservative 25; Mismatches 95; Indels 63; Gaps 11;

QY 1 MKFTITSSALALALGTEMAAARSKLYGCGGKDMNPGTCESGTCVSNQYSGC 60
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 1 MGLKGTAAALTLTSLA---AAASAGCVLEQGGIGFDGTCSEGLMCMKMGPIYSQC 57
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 61 LA-PESNGKSSSECSKLYGCGGKDMNPGTCESGTCVSNQYSGCLAPESNKNKTS 119
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 58 RAMPGMGQ---VAPYGGCGGMNYSKTMSPGPKCVLELNFPSQC---DLANKSPV 108
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 120 SAHKTTTTTAPAKETTTAKASNSNSGKSYSVSGGAGNGVTRYMCKC-----A 172
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 109 APKPVAPTSP-----GPAQVCGEYVAACGEMFWGA---KCCFGLVCYET 152
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 173 SCGMPGKANVSFVKSCKNDGVTALSDSNVSGCGNGANSY---MCNDNQPAVNDNLAY 228
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 153 SGKMGSGCCRAPPP-----KMEVGKRYAC-GGKMGSTMC-----VG 189
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 229 GFAAAAISSGESRMC 244
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 190 GYKCMALISEGSMYKQC 205
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 6
ID GUX2 TRIRE STANDARD; PRT; 471 AA.
AC P07987;
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellulohydrolase II)
DE (CBHI) (1,4-beta-cellobiohydrolase).
GN CBH2.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=VTT-D-80133;
RX MEDLINE=87248061; PubMed=3596237;
RA Teeri T.T., Lehenavaara P., Kauppinen S., Salovuori I., Knowles J.;
RT "Homologous domains in Trichoderma reesei cellulytic enzymes: gene
   sequence and expression of cellobiohydrolase II.",
   Gene 51:43-52 (1987).
RL 12
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=QM9414 / Rut C-30;
RA Chen C.W., Gritzali M., Stefford D.W.;
RT "Nucleotide sequence and deduced primary structure of
   cellobiohydrolase II from Trichoderma reesei.",
   Biotechnology 5:274-278 (1987).
RL 13
RP [3]
RA Paegerstam L.G., Petersson L.G.;

```

```

RT "The 1,4-beta-glucan cellobiohydrolases of Trichoderma reesei QM
   9414.",
   FEBS Lett. 119:97-100(1980).
RL 14
RN 14
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=90333255; PubMed=2377893;
RA Rouvinen J., Bergfors T., Teeri T.T., Knowles J.K.C., Jones T.A.;
RT "Three-dimensional structure of cellobiohydrolase II from Trichoderma
   reesei.",
   Science 249:380-386 (1990).
RN 15
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=97029636; PubMed=8875646;
RA Koivula A., Reinikainen T., Ruohonen L., Valkeajarvi A., Rouvinen J.,
   Claessens M., Teeman O., Kleywegt G.J., Szardenings M., Rouvinen J.,
   Jones T.A., Teeri T.T.;
RA "The active site of Trichoderma reesei cellobiohydrolase II: the role
   of tyrosine 169.",
   Protein Eng. 9:691-699(1996).
RL 16
RN 16
RP FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC -1-
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIHYDROLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: T. RESEI PRODUCES TWO DIFFERENT
CC EXOCELLULOBIHYDROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC EMBL; M16190; AAA4210.1; -
CC EMBL; M55080; AAA72922.1; -
CC PIR; A26472; A26472.
CC PIR; A26160; A26160.
CC PDB; 3CBH; 15-JAN-91.
CC PDB; 1CB2; 19-MAR-93.
CC InterPro; IPR000254; CBD_fungal.
CC InterPro; IPR01524; GH_6.
CC Pfam; PF00734; CBM_1; 1.
CC Pfam; PF01341; GHYDRLASE6.
CC PRINTS; PR00733; GHYDRLASE6.
CC ProDom; PD001821; GH_6; 1.
CC ProDom; PD003733; GH_6; 1.
CC SMART; SM00236; FCBP; 1.
CC PROSITE; PS00562; CBD_FUNGAL; 1.
CC PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
CC PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
   3D-structure.
KM 3D-structure. 1 24
FT SIGNAL 25 471 EXOGLUCANASE II.
FT CHAIN 25 65 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 25 106 LINKER.
FT DOMAIN 66 107 CATALYTIC.
FT DOMAIN 107 199
FT ACT_SITE 199 199
FT ACT_SITE 245 245 PROTON DONOR.
FT ACT_SITE 245 425 NUCLEOPHILE.
FT ACT_SITE 425 51 BY SIMILARITY.
FT DISULFID 34 51

```

```

FT DISULFID 45 61 BY SIMILARITY.
FT DISULFID 200 259
FT DISULFID 392 439
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 O-LINKED (MAN. . .)
FT CARBOHYD 121 121 O-LINKED (MAN. . .)
FT CARBOHYD 130 130 O-LINKED (MAN. . .)
FT CARBOHYD 133 133 O-LINKED (MAN. . .)
FT CARBOHYD 134 134 O-LINKED (MAN. . .)
FT CARBOHYD 139 139 O-LINKED (MAN. . .)
FT CARBOHYD 313 313 N-LINKED (GLCNAC. . .)
FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .)
FT MUTAGEN 199 199 D->A: 20% OF WILD-TYPE ACTIVITY.
FT MUTAGEN 245 / 245 P->A: NO MEASURABLE ACTIVITY.
FT CONFLICT 359 359 P-> R (IN REF. 2).
FT CONFLICT 449 449 P-> A (IN REF. 2).
SQ SEQUENCE 471 AA; 49653 MW; C4711BC35B1BD8 CRC64;

Query Match 11.3%; Score 227.5; DB 1; Length 471;
Best Local Similarity 22.7%; Pred. No. 2.3e-08;
Matches 113; Conservative 38; Mismatches 159; Indels 187; Gaps 22;

QY 4 ITTSSALLALALCTEMASAKSKLYGCGCGKDMNGPTCCSGSTCKVSNDRYSQCLAP 63
DB 5 ILTLATLATLATAVPLEERQACSSVMGCGGKDMNGPTCCSGSTCKVSNDRYSQCLAP 64
QY 64 ESNMKSSECKLYGCGCGKDMNGPTCCSGS-----TCRY----- 99
DB 65 AASSSSSTRASTSTRV-----SPTTSRSSATPPPGSTTTFVPNGSGTATYSGNPFV 118
QY 100 -----SNDY-----SCLAPESNGKTSBSA-----HKTITTAPEKITTT 136
DB 119 GVTPMANAYVASEVSLAPSLTGAMATAAAVAKVPSFMMLLTLLDTKPLMEQTLADIRT 178
QY 137 TAKASNSN-----SGKTSIVSGCAGS-----NGVTRTWDC 169
DB 179 ANK--NGGNVAGQFVVYDLPDRDCAALASNEYSIADGKAVKXNYIDTIRQIVVEYSDI 236
QY 170 -----CKASGWPGRKANVSFVSCNKDGTALSDSNVSGCGNGSYMND 216
DB 237 RTLLVIEPDSLANTVTNIGTKCANAGSAYIECTINVAITLNPVANYTLDAHGHWLG- 295
QY 217 NQPMVNDNLAYGFPAALAISSGSGSR-----WCCSCFELTFTSTVAGKRM 262
DB 296 ---WPANDPAAQLFANYKYKASSRRLAGLATVANYANGNITSPSPYTGNAVYMEKL 352
QY 263 VIO-----VTNTG-----GDLSSSTGAHFDLQPGGSGVIFNGCSKRWG----- 301
DB 353 YIHAIGPLLANHGWSNAFFITDQGRSG-----KQPTG-----QQQWDMCNVIGTG 398
QY 302 -----APRDG-----WGSRYGSISSADCS-----SLPSAL-----QAGCKRKN 336
DB 399 FGIRPSANTGDSLLDSFVWVRGECDDGTSASAPRFDSCALPDALQAPADAGA----- 453
QY 337 WFK-----NADNPS 345
DB 454 WFOAVFVQQLTNA-NPS 469

RESULT 7
GUX3_AGABI STANDARD; PRT; 438 AA.
AC P49075;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE Exoglucanase 3 precursor (EC 3.2.1.91) (Exocellulohydrolase 3)
DE (1,4-beta-cellulohydrolase 3).
GN CEL3.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OC NCBI_TaxID=5341;

```

```

RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 255-277 AND 331-351.
RC STRAIN=D649;
RX MEDLINE=94368092; PubMed=8085821;
RA Chow C.-M., Yague E., Raguz S., Wood D.A., Thurston C.F.;
RT "The cel3 gene of Agaricus bisporus codes for a modular cellulase and
RL is transcriptionally regulated by the carbon source.";
RN [2]
RP Appl. Environ. Microbiol. 60:2779-2785(1994).
RC TISSUE=Mycelium;
RX MEDLINE=96269930; PubMed=8662210;
RA Yague E., Chow C.-M., Challen M.P., Thurston C.F.;
RT "Correlation of exons with functional domains and folding regions in
RL a cellulase from Agaricus bisporus.";
RN [3]
RP Curr. Genet. 30:56-61(1996).
CC -1- FUNCTION: SHOWS ENZYMIC ACTIVITY TOWARDS CRYSTALLINE CELLULOSE.
CC AT LONG REACTION TIMES. IT IS ALSO ABLE TO DEGRADE CARBOXYMETHYL
CC CELLULOSE AND BARLEY B-GLUCAN.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC HYDROLASES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24519; AAA50607.1; -
DR EMBL; L24520; AAA50608.1; -
DR EMBL; Z34007; CA83971.1; -
DR HSSP; P00725; 1A26.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001524; GH_6.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF01341; Glyco_hydro_6; 1.
DR PRINTS; PR00733; GLHYDRLA66.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD003733; GH_6; 1.
DR SMART; SM00236; FCBF; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 20
FT CHAIN 21 438 EXOGLUCANASE 3.
FT DOMAIN 21 59 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 60 87 LINKER.
FT DOMAIN 88 438 CATALYTIC.
FT ACT_SITE 215 215 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 393 393 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 28 45 BY SIMILARITY.
FT DISULFID 39 55 BY SIMILARITY.
FT DISULFID 170 229 BY SIMILARITY.
FT DISULFID 360 407 BY SIMILARITY.
FT VARIANT 133 133 V->T.
FT VARIANT 152 152 R->Q.
FT VARIANT 244 244 V->I.
FT VARIANT 248 248 N->S.
FT VARIANT 398 398 N->D.
SQ SEQUENCE 438 AA; 46209 MW; 002C973544893794 CRC64;

Query Match 8.8%; Score 177.5; DB 1; Length 438;
Best Local Similarity 24.7%; Pred. No. 4.4e-05;
Matches 107; Conservative 41; Mismatches 151; Indels 135; Gaps 21;

QY 9 SALLALAL--GTEMASAKSKLYGCGCGKDMNGPTCCSGSTCKVSNDRYSQCL----- 61

```

```

Db      5 AALLALALVPEFVQAQ-----SPWVGCGGANGMTGPTTCAGSGSTCVKQNDPFSQCLPNNQ 60
Qy      62 APESNGNKSSECSKLYGQCGKDMNNGPTCC-----SGSTCKYS--NDYSGCLAPESNG 114
Db      61 APPSTTTPGTTPPATTTSGG---TGPTSGAGNPYTGKTVLWSPRYADEVAQAADISNP 117
Qy      115 NNTSSAH--KTTT-----TTAPAKEITTTAKASNSN-----SSG 148
Db      118 SLATKAASVAKIPTVWPTVAKVVDLGVIADASKQVQVIVYDLPDDCAALASNG 177
Qy      149 KXSVSGAS-----GNGVTT-RYMDCCASGSCWP 177
Db      178 EESLNDGINKYKANYDQIAAQIKOPDPVSVVAVIEPDLANLVNINLVOKANQASAY- 236
Qy      178 GKAANSSPYKSCNKGVTALSDSNVSGCGNGSNVWCNDNQWAVNDLALGFAAALASG 237
Db      237 -KEGYIYAVQKLNAGVTVYIDAG-HAGWLG-----WPAHLSPPAQLFQIYRD 283
Qy      238 GGESRWCCSCFELTSTSVAGKKVY-----IQVTVGCD-----LGSSTG-- 278
Db      284 AASPR-----NLRGATVAVANFALRASPPDITQGSNTYDEHYIEALAPMLSNAGFP 337
Qy      279 AHFDLQMPGGVGIFFNGCSKQWAPNDGSGRYGSISSADCSLPSALQACCKRFNWF 338
Db      338 AHFIVDQGRGV---QNIQDQWDMCNVKGAGFGQPTNTGSSLDIAIV-----WV 386
Qy      339 K-----NADNPS 345
Db      387 KPGECGDTSDNS 400

```

```

RESULT 8
GUN2_TRIPE STANDARD; PRT; 418 AA.
ID GUN2_TRIPE
AC P07982;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE Endoglucanase EG-II precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN EGL2 OR EGLII.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
ON NCBI_TaxID=51453;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VIT-D-80133;
RC MEDLINE=88255850; PubMed=3384334;
RA Saloheimo M., Lehtovaara P., Penttilae M., Teeri T.T., Staahlberg J.,
RA Johansson G., Pettersson G., Claessens M., Tomme P., Knowles J.K.C.;
RT "EGLII, a new endoglucanase from Trichoderma reesei: the
RT characterization of both gene and enzyme.";
RT Gene 63:11-21 (1988).
RN [2]
RP ACTIVE SITE GLU-350.
RX MEDLINE=9311031; PubMed=8093602;
RA Macaron R., van Beeunen J., Henriessat B., de la Mata I.,
RA Claessens M.;
RT "Identification of an essential glutamate residue in the active site
RT of endoglucanase III from Trichoderma reesei.";
RL FEBS Lett. 316:137-140 (1993).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOLUCANASES WHICH CUT INTERNAL BETA-1,4-GLYCOSIDIC BONDS;
CC (2) EXOCELLULOBIODOLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLULO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL

```

```

CC      CC HYDROLASES).
CC      CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC      CC -1- CAUTION: WAS ORIGINALLY CALLED ENDOGLUCANASE EG-II.
CC      CC
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      CC the European Bioinformatics Institute. There are no restrictions on its
CC      CC use by non-profit institutions as long as its content is in no way
CC      CC modified and this statement is not removed. Usage by and for commercial
CC      CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC or send an email to license@isb-sib.ch).
CC      CC
CC      CC EMBL, M19373; AAA34213.1; -.
CC      CC PIR, S28372; S28372.
CC      CC HSSP, P00725; ZCBH.
CC      CC InterPro: IPR000254; CBD_fungal.
CC      CC InterPro: IPR001547; GH 5.
CC      CC Pfam, PF00734; CBM_1; 1.
CC      CC Pfam, PF00734; CBM_1; 1.
CC      CC ProDom, PD001821; CBD_fungal; 1.
CC      CC SMART, SM00236; fcbd; 1.
CC      CC PROSITE, PS00562; CBD_FUNGAL; 1.
CC      CC PROSITE, PS00569; GLYCOSYL_HYDROL_F5; 1.
CC      CC Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
CC      CC
CC      CC SIGNAL
CC      CC 1
CC      CC 21
CC      CC 22 418 ENDOGLUCANASE EG-II.
CC      CC 22 57 CELLULOSE-BINDING (BY SIMILARITY).
CC      CC 58 91 LINKER.
CC      CC 92 418 CATALYTIC.
CC      CC 92 418 PYROLIDONE CARBOXYLIC ACID.
CC      CC 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      CC 124 124 BY SIMILARITY.
CC      CC 29 46 BY SIMILARITY.
CC      CC 40 56 PROTON DONOR (BY SIMILARITY).
CC      CC 239 239 NUCLEOPHILE.
CC      CC ACT SITE 350
CC      CC FT ACT SITE 350
CC      CC FT ACT SITE 350
CC      CC SQ SEQUENCE 418 AA; 44227 MW; 26A492D55237A49B CRC64;

```

```

Query Match 8.3%; Score 167.5; DB 1; Length 418;
Best Local Similarity 25.7%; Pred. No. 0.00019;
Matches 62; Conservative 28; Mismatches 66; Indels 85; Gaps 12;
Qy      76 LYGGCGGKDMNNGPTCCSGSTCKVSNQCLAPESNGNKTSESAHKTITTTAPAKEIT 135
Db      25 WMGCGGIGGMSGPNNCARGSACSTLNRYAQCI-----PGATITTSRTPPGPT 74
Qy      136 TTAASNSNSGKYSIVSGASGNTTRYWDCCKASGMPG---KANVSPYKS----- 188
Db      75 TTRATSTSSSTPP---TSSGVRPAGVNIAGFD---FGCTTDTGCVTSKYVPLKNTGS 128
Qy      189 -----CNKQVTA-----LSDSNVSGCGNCGSNVWCNDNQWAVANDNL 226
Db      129 NNPYDGIQOMQHFVNEEDMTFLRPLVGQVYVNNNL-----GQNL-----DSTISKIDOL 179
Qy      227 AYGFAPAAISGGGSRWCCSCFELTFTSTSVAGKKVYQVTVN---TGGLGSS--TGAH 280
Db      180 VQG-----CLSL-----GAYCIVDHNVARMNGGIIIGCGGPTNAQ 214
Qy      281 F 281
Db      215 F 215

```

```

RESULT 9
GUX1_ASPAC STANDARD; PRT; 540 AA.
ID GUX1_ASPAC
AC 059843;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulohydrolase I) (1,4-
DE beta-cellulohydrolase I) (Beta-glucanocellulohydrolase I).
GN CBHI.
OS Aspergillus aculeatus.

```



```

CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxId=5053;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F-50;
RA Takada G., Kawaguchi T., Sumitani J., Arai M.;
RT "Cloning, nucleotide sequence, and transcriptional analysis of
RT Aspergillus aculeatus No.F-50 cellobiohydrolase I (cbhI) gene.";
RL J. Ferment. Bioeng. 85:1-9(1998).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLYCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOGLYCOSYLASES WHICH CUT THE DISACCHARIDE CELLOBIOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellobiose, releasing cellobiose from the non-
CC reducing end of the chains.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY C (FAMILY 7 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB002821; BAA25183.1; -.
DR HSSP; P00725; 1A26.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR001722; GH_7.
DR Pfam; PF00734; CBM_1; 1_1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; GH_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00563; CBD_FUNGAL; FALSE NEG.
KW Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 540
FT DOMAIN 23 459
FT DOMAIN 506 505
FT DOMAIN 513
FT ACT_SITE 234 234
FT ACT_SITE 239 239
FT DISULFID 513 529
FT DISULFID 524 539
FT CARBOHYD 434 434
FT CARBOHYD 458 458
SQ SEQUENCE 540 AA; 57059 MW; 002D7FD2BDF194D0 CRC64;
Query Match 7.8%; Score 158.5; DB 1; Length 540;
Best Local Similarity 41.3%; Pred. No. 0.00096;
Matches 31; Conservative 9; Mismatches 34; Indels 1; Gaps 1.
OY 34 GCKDMNGPTCCESGTCVKVSNNDYISQCLAPSNKMSKSECSLYQCGCKDMNGPTCCS 93
Db 467 GGSSSSTLTLLTKKASTSTSSKTYTTTSKTSSTSSSTVAOLVCGCGGCGMTGPTCAS 526
OY 94 GSTCKVSNDDYSQCL 108
Db 527 G-TCTKONDYISQCL 540
RESULT 10
ID GUNS_TRIRE STANDARD; PRT; 242 AA.
AC PA3317;

```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DB Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DB (Cellulase V) (EG V).
GN EGL5.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Peizizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_taxid=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM9414 / Rut C-30;
RX MEDLINE=95075308; Pubmed=7984103;
RA Satoheimo A., Hentrisat B., Hoffren A.-M., Tejeman O., Penttilae M.;
RT "A novel, small endoglucanase gene, egl5, from Trichoderma reesei
RT isolated by expression in yeast."
RL Mol. Microbiol. 13:219-228(1994).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL
CC HYDROLASSES).
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DR EMBL; Z33381; CAA83846.1; -.
DR HSSP; P00725; 2CBB.
DR InterPro; IPR000254; CBD_fungal.
DR InterPro; IPR000334; GH_45.
DR Pfam; PF00734; CBM_1; 1_45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR ProDom; PD001821; CBD_fungal; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR PROSITE; PS00842; EXPANSIN_EG45; 1.
DR PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 17
FT CHAIN 18 242
FT ACT_SITE 27 27
FT ACT_SITE 134 134
FT DOMAIN 18 182
FT DOMAIN 183 205
FT DOMAIN 206 242
FT CARBOHYD 182 182
FT DISULFID 213 230
FT DISULFID 224 240
FT SEQUENCE 242 AA; 24411 MW; CC033FC51326C71D CRC64;
Query Match 7.6%; Score 154; DB 1; Length 242;
Best Local Similarity 40.5%; Pred. No. 0.0009;
Matches 32; Conservative 10; Mismatches 29; Indels 8; Gaps 2;
QY 33 CGGKDMNCGPTCC--ESGSKTVKNDYISQCLAPESNKNKSECKLYGCGGCKDMNPT 89
DB 168 CVGGQEDNDPFTVLNDGTSPPSSPATSSPPSGGQQT---LVGCGCGAGMTGPT 222
QY 90 CCGSGSTCKVSNDDYSOCL 108
DB 223 TCQAPGCTKVQNMWISQCL 241
RESULT 11
GUNB_GUNB_FUSOX STANDARD. PRT; 462 AA.
AC P46236;

```

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Putative endoglucanase type B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).

OS Fusarium oxysporum.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Microsporici Hypocreales; Fusarium.

OX NCBI_TaxID=5507;

XX MEDLINE=95047531; PubMed=7959045;

RA Shepherd P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C., Hagen F.S., Uphall A., McKnight G.L., O'Hara P.J.;

RT "The use of conserved cellulase family-specific sequences to clone cellulase homologue cDNAs from *Fusarium oxysporum*."

RL Gene 150:163-167(1994).

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose.

CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY B (FAMILY 6 OF GLYCOSYL HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; I29377; AAA65585.1; -

DR HSSP; P07987; 1CB2.

DR InterPro; IPR000254; CBD_fungal.

DR InterPro; IPR001524; GH_6.

DR Pfam; PF00734; CBM_1; 1.

DR Pfam; PF01341; Glyco_hydro_6; 1.

DR PRINTS; PR00733; GLHYDRLASE6.

DR ProDom; PD001821; CBD_fungal; 1.

DR ProDom; PD003733; GH_6; 1.

DR SMART; SM00236; ECBD; 1.

DR PROSITE; PS00562; CBD_FUNGAL; 1.

DR PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.

DR PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.

DR Cellulose degradation; Hydrolyase; Glycosidase; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 462

FT DOMAIN 17 65

FT DOMAIN 66 99

FT DOMAIN 100 462

FT ACT_SITE 190 236

FT ACT_SITE 236 236

FT ACT_SITE 416 416

FT DISULFID 33 50

FT DISULFID 44 60

FT DISULFID 191 250

FT DISULFID 383 430

FT CARBOHYD 37 37

FT CARBOHYD 223 223

FT CARBOHYD 272 272

FT CARBOHYD 317 317

FT CARBOHYD 462 AA; 49207 MW; E25B25B828B637F CRC64;

SEQUENCE

Query Match 7.5%; Score 152; DB 1; Length 462;

Best Local Similarity 34.8%; Pred. No. 0.0022;

Matches 31; Conservative 14; Mismatches 36; Indels 8; Gaps 2;

QY 73 CSK-LYGGCGGDMNGPTCCSGSTCKVSNDRYSGCLAPENGNKTSNAKTTTTPA 131

DB 25 CSNGWAGCGGDMNGPTCCSGSTCKVSNDRYSGCLAPENGNKTSNAKTTTTPA 80

QY 132 KEITTAASNSNSGKGYIVSGASGN 160

Db 81 ---KTATGSSSTTAGSVTSAPPAASDN 106

RESULT 12

AC JAG1_HUMAN STANDARD; PRT; 1218 AA.

AC P78504; O15122; O14902; Q15816;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Jagged 1 precursor (Jagged1) (Jag1).

GN JAG1.

OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

XX MEDLINE=97422615; PubMed=9268641;

RA Oda T., Elkanloun A.G., Welter P.S., Chandrasekharappa S.C.;

RT "Identification and cloning of the human homolog (JAG1) of the rat Jagged1 gene from the Alagille syndrome critical region at 20p12."

RL Genomics 43:376-379(1997).

CC [2]

CC SEQUENCE FROM N.A., AND VARIANT AGS CVS-184.

CC TISSUE=Bone marrow;

CC MEDLINE=97351506; PubMed=9207788;

RA Li L., Krantz I.D., Deng Y., Genin A., Banta A.B., Collins C.C., Qi M., Trask B.J., Kuo W.L., Cochran J., Costa T., Pierpont M.E.M., Rand E.B., Piccoli D.A., Hood L., Spinner N.B.;

RT "Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1."

RL Nat. Genet. 16:243-251(1997).

CC [3]

CC SEQUENCE FROM N.A., AND FUNCTION.

CC MEDLINE=98122342; PubMed=9462510;

RA Li L., Milner L.A., Deng Y., Iwata M., Banta A.B., Graf L., Marcovina S., Friedman C., Trask B.J., Hood L., Torok-Storb B.;

RT "The human homolog of rat Jagged1 expressed by marrow stroma inhibits differentiation of 32D cells through interaction with Notch1."

RL Immunity 8:43-55(1998).

CC [4]

CC SEQUENCE FROM N.A.

CC TISSUE=Cervical carcinoma;

CC MEDLINE=99262417; PubMed=10329625;

RA Bash J., Zong W.-X., Banga S., Rivera A., Ballard D.W., Ron Y., Gelinas C.;

RT "Rel/NF-kappaB can trigger the Notch signaling pathway by inducing the expression of Jagged1, a ligand for Notch receptors."

RL EMBO J. 18:2803-2811(1999).

CC [5]

CC SEQUENCE FROM N.A.

CC MEDLINE=99180765; PubMed=10079256;

RA Gray G.E., Mann R.J., Mitsiadis E., Henrique D., Carcangiu M.L., Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;

RT "Human ligands of the Notch receptor."

RL Am. J. Pathol. 154:785-794(1999).

CC [6]

CC SEQUENCE FROM N.A.

CC MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Copley V.E., Collier R.E., Connor R.E., Cotby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Gathani D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond D., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,


```

Db 538 AQCYNRADSYFCKC--PEYEGKNSHLKDHCRTPCEVIVSCYVAMASNTPEGVX-- 593
Oy 153 VSGASGNGVTRRYWDCCASCSWPGKANVSPV---SCKKQVTA-----LSDSNVQ 203
Db 594 ----ISSN-----VCGPHGKCKSOGGKFTCDCKN-GPTGYCHENINDCSN 636
Oy 204 SGCNGG-----NSYMCNDNCPWA-----VNDLAVGFMAAALSGGESR-----WCC 245
Db 637 PCRNNGTCTIDGVNSYKICISDMEGAYCETINND-----CSQNPCHNGGTCRDVLVNDPYC 691
Oy 246 SCFELFTSTSVAKKMYIQVNTNGDLSSTGAHFDLQMPG----- 287
Db 692 DCKNKGKGTCHSRPSQCEATCNNGGTCYDGDAFKCMCPGMEGTTCNLARNSCLPN 751
Oy 288 ----GGVGIFNG-----CSKMGAPNDGWSRGYGISASLSCSLPSLQAGCKMRNW 337
Db 752 PCHNGGTCVNGNESFTVCCKEGMEGPICA-----QNTNDCSPHPCYNSGTCVGDGNW 803
Oy 338 FK 339
Db 804 YR 805

RESULT 13
JAG1_MOUSE STANDARD; PRT; 1218 AA.
ID 09QXK0;
AC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Jaged 1 precursor (Jaged1).
GN JAG1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A., AND RECEPTOR INTERACTION.
RC STRAIN=Swiss Webster / NIH;
RX MEDLINE=20020271; PubMed=10551863;
RA Shimizu K., Chiba S., Kumano K., Hosoya N., Takahashi T., Kanda Y.,
RA Hamada Y., Yazaki Y., Hirai H.;
RT "Mouse Jaged1 physically interacts with Notch2 and other Notch
RT receptors: assessment by quantitative methods.";
RL J. Biol. Chem. 274:32961-32969(1999).
RN [2]
RP TISSUE SPECIFICITY
RX MEDLINE=20025753; PubMed=10556292;
RA Loomes K.M., Underkofler L.A., Morabito J., Gottlieb S.,
RA Piccoli D.A., Spinner N.B., Baldwin H.S., Oakey R.J.;
RT "The expression of Jaged1 in the developing mammalian heart
RT correlates with cardiovascular disease in Alagille syndrome.";
RL Hum. Mol. Genet. 8:2443-2449(1999).
CC -1- FUNCTION: Ligand for multiple Notch receptors and involved in the
CC mediation of Notch signaling. May be involved in cell-fate
CC decisions during hematopoiesis. Seems to be involved in early
CC and late stages of mammalian cardiovascular development. Inhibits
CC myoblast differentiation (By similarity). May regulate fibroblast
CC growth factor-induced angiogenesis.
CC -1- SUBUNIT: Interacts with NOTCH1, NOTCH2 AND NOTCH3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Widely expressed in many tissues, with highest
CC expression in brain, heart, muscle and thymus.
CC -1- DEVELOPMENTAL STAGE: At 8.75-9.75 d.p.c. expression was detected
CC in structures that include those destined to contribute to the
CC cardiovascular system of the adult heart. Expression was also
CC detected in the mesencephalon and rhombencephalon.
CC -1- DOMAIN: The DSL domain is indispensable and sufficient for binding
CC to NOTCH2.
CC -1- SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DSL DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AF171092; AAF15505.1; -.
CC HSSP; P00740; 1EDM.
CC MED; MG1.1095416; JAG1.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR001774; DSL.
CC InterPro; IPR000561; EGF-Like.
CC InterPro; IPR000742; EGF 2.
CC InterPro; IPR001881; EGF Ca.
CC InterPro; IPR001438; EGF_II.
CC InterPro; IPR002049; laminin_EGF.
CC InterPro; IPR010071; WVF_C.
CC Pfam; PF01414; DSL; 1.
CC Pfam; PF00008; EGF; 14.
CC PRINTS; PR00010; EGFBLD.
CC PRINTS; PR00011; EGFAMININ.
CC SMART; SM00051; DSL; 1.
CC SMART; SM00179; EGF_CA; 10.
CC SMART; SM00001; EGF_Like; 6.
CC SMART; SM00214; WVF; 1.
CC PROSITE; PS00010; ASX_HYDROXYL; 10.
CC PROSITE; PS00022; EGF_1; 16.
CC PROSITE; PS01186; EGF_2; 12.
CC PROSITE; PS01187; EGF_CA; 8.
CC KMW Calcium-binding; EGF-like domain; Glycoprotein; Developmental protein;
CC Repeat; Transmembrane; Signal.
CC FT SIGNAL 1 33
CC FT CHAIN 34 1218
CC FT DOMAIN 34 1067
CC FT TRANSMEM 1068 1093
CC FT DOMAIN 1094 1218
CC FT DOMAIN 167 229
CC FT DOMAIN 230 266
CC FT DOMAIN 296 334
CC FT DOMAIN 336 372
CC FT DOMAIN 374 410
CC FT DOMAIN 412 448
CC FT DOMAIN 450 485
CC FT DOMAIN 487 523
CC FT DOMAIN 525 561
CC FT DOMAIN 574 627
CC FT DOMAIN 629 665
CC FT DOMAIN 667 703
CC FT DOMAIN 705 741
CC FT DOMAIN 744 780
CC FT DOMAIN 782 816
CC FT DOMAIN 820 856
CC FT DISULFID 300 312
CC FT DISULFID 306 322
CC FT DISULFID 324 333
CC FT DISULFID 340 351
CC FT DISULFID 345 360
CC FT DISULFID 362 371
CC FT DISULFID 378 389
CC FT DISULFID 383 398
CC FT DISULFID 400 409
CC FT DISULFID 416 427
CC FT DISULFID 421 436
CC FT DISULFID 438 447
CC FT DISULFID 454 464
CC FT DISULFID 458 473
CC FT DISULFID 475 484
CC FT DISULFID 491 502
CC FT DISULFID 496 511
CC FT DISULFID 513 522
CC FT DISULFID 529 540

```

```

FT DISULFID 534 549 BY SIMILARITY.
FT DISULFID 551 560 BY SIMILARITY.
FT DISULFID 578 605 BY SIMILARITY.
FT DISULFID 599 615 BY SIMILARITY.
FT DISULFID 617 626 BY SIMILARITY.
FT DISULFID 633 644 BY SIMILARITY.
FT DISULFID 655 664 BY SIMILARITY.
FT DISULFID 671 682 BY SIMILARITY.
FT DISULFID 676 691 BY SIMILARITY.
FT DISULFID 693 702 BY SIMILARITY.
FT DISULFID 709 720 BY SIMILARITY.
FT DISULFID 714 729 BY SIMILARITY.
FT DISULFID 731 740 BY SIMILARITY.
FT DISULFID 748 759 BY SIMILARITY.
FT DISULFID 753 768 BY SIMILARITY.
FT DISULFID 770 779 BY SIMILARITY.
FT DISULFID 786 797 BY SIMILARITY.
FT DISULFID 791 806 BY SIMILARITY.
FT DISULFID 808 817 BY SIMILARITY.
FT DISULFID 824 835 BY SIMILARITY.
FT DISULFID 829 844 BY SIMILARITY.
FT DISULFID 846 855 BY SIMILARITY.
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 559 559 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1064 1064 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1218 AA; 134163 MW; 77739F89288B793C CRC64;

```

Query Match 7.3%; Score 147.5; DB 1; Length 1218;
 Best Local Similarity 20.9%; Pred. No. 0.011;
 Matches 89; Conservative 39; Mismatches 147; Indels 151; Gaps 24;

```

QY 22 SAACSKL-----YQCGGKDMNGPTC-----CISGSKT-VSNDYYSQCLAPESN 66
DB 423 NARSKKNLIASYYCDC-LPGWMGQNCNDINDCIQCONDASCRDLVNGY--RCICPP-- 477
QY 67 GNKSESEKLYGQC-----CGK-----DMNGPTCESE 94
DB 478 GYAGDHCRDRIDECSNPCLANGHCONEINRQCLCPFGSGNLCOLDIDICEPPCQNG 537
QY 95 STC-KVSNDDYYSQCLAPES-NGNKTSESAHKTITTTAPAKEITTTAKASNSNSGKY-- 150
DB 538 AQCVRASADYFCCK--PEDYEGKNSHLKDHCRITTCETVIDSCTVAMASNDTPBEGVRIYS 595
QY 151 SIVSG-----GASGNVYTRRYMDCCKASCSWPKAANVSSPVKSCNKGVTALSDSNVOS 204
DB 596 SNVCGPHGKCKSOSGKFTCC--DCKNKPRTGYCHEININDCSNPKCKNGGTCTI----- 645
QY 205 GCNGNSYMCNDNQPMW-----VND-----NLAYGFAAAT-----SGGSESRWCC 245
DB 646 --DGNVSTKICICSDMEGAHCENNNINDCSQNPCHYGICRDI-VNDPYCDCKRGMKGKTC 703
QY 246 S-----CFELTFTSTVSAAKKNVIOVTNTGGDLGSGTGAHFDLQMG----- 287
DB 704 SRDSCDEAT-----CNGGTCYDEVDI-FKCMCGGMEGTTCNTIARRSS 747
QY 288 -----GGVGIFNG-----CSKQWGAIPNDGMSRIGGISASADCSLPALQAGCKW 333
DB 748 CLPNCHNGGTCVNVGDSFTVCYCKEGWGPFI-----CTGYTNDCSPHPCYNSTGTCVD 799
QY 334 RPNWFK 339
DB 800 GDNWTR 805

```

RESULT 14
 Y136_TREPA

```

ID Y136 TREPA STANDARD; PRT; 485 AA.
AC O83172;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical Lipoprotein TP0136 precursor.
GN TP0136.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RA [1]
RA SEQUENCE FROM N. A.
RA STRAIN=Nichols;
RX MEDLINE=9833270; PubMed=965876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA "Complete genome sequence of Treponema pallidum, the syphilis
RA spirochete."
RL Science 281:375-388(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AE001199; AAC65137.1; ALT_INIT.
CC TIGR; TP0136; -.
CC KW Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
CC FT SIGNAL 1 23 POTENTIAL.
CC FT CHAIN 24 485 HYPOTHETICAL LIPOPROTEIN TP0136.
CC FT LIPID 24 24 N-ACYL DIGLYCERIDE (POTENTIAL).
CC FT DOMAIN 164 178 GLY/SER-RICH.
CC FT DOMAIN 196 210 GLY/SER-RICH.
CC FT DOMAIN 253 267 GLY/SER-RICH.
CC FT DOMAIN 318 327 POLY-SER.
CC FT DOMAIN 444 447 POLY-SER.
CC FT DOMAIN 485 AA; 48984 MW; C7A4CEDEC7DC5CED CRC64;
SQ SEQUENCE

```

Query Match 7.2%; Score 145; DB 1; Length 485;
 Best Local Similarity 21.7%; Pred. No. 0.0068;
 Matches 75; Conservative 43; Mismatches 131; Indels 96; Gaps 12;

```

QY 27 SKLY-----GQCGGKDMNGPTCESGTCVSNDDYYSQCLAPESNGKSESEKLYGQCGSK 83
DB 56 SKLYATNRRLMEKELNGSGWQKYSVSPDSDKWNISATDGTFF----- 103
QY 84 DMNGPTCESGTCVSNDDYYSQCLAPESNGKTSSEAHKTITTTAPAKEITTTAKASNS 143
DB 104 -----LACVPTGTVYKHCY-----NGAGSS-----STGTASPESTE-TCSGHATLV 143
QY 144 SNSSGKTSIVSGASGNGVTRRYMDCCASCSWPKAANVSSPVKSCNKGVTALSDSNVQ 203
DB 144 GGTSKPFLVLPQGTGNNG-----NCGCGGGGSSSSSSSCTIHWLPQGTGNNGN 194
QY 204 GCNGNSYMCNDNQPMWAVNINLAYGFAAAAISGGSESRWCCCFELTFTSTVS-----AG 259
DB 195 CGCGGG-----GGSSSSSSSSSCTIHWLVNTDEQFLMG 227
QY 260 KKNVIOVTNTGGDLGSGTGAHFDLQMPGGVGIVFNGCSKQWGAIPNDGMSRIGGISASD 319
DB 228 EGYVVTGHLVTKNGSSSAG--PAQCPGGGG-----GSSGCGGSSSEYTVASCSF 276

```

QY 320 CSSLSALQAGCKMFPNMFKNADNSMTYKEYTCPEKITAKTCCS 364
 Db 277 STPIIASVSDGC---YHYI-----LTREKVCYCRKODTASSAAS 311

RESULT 15
 JAG1_BRARE STANDARD; PRT; 1242 AA.
 ID JAG1_BRARE
 AC Q90Y57;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Jagged 1 precursor (Jagged1).
 GN JAG1.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Oda T., Chandrasekharappa S.C.;
 RT "Isolation, characterization and expression analysis of Zebrafish
 RT Jagged genes.";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Ligand for multiple Notch receptors and involved in the
 CC mediation of Notch signaling (By similarity). Seems to be involved
 CC in cell-fate decisions.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
 CC -1- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 DSL DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; AF229448; AL08213.1; -
 CC DR ZFIN; ZDB-GENE-011128-2; JAG1.
 CC DR InterPro; IPR000152; Asx_hydroxyl.
 CC DR InterPro; IPR001774; DSL.
 CC DR InterPro; IPR000561; EGF-like.
 CC DR InterPro; IPR000742; EGF 2.
 CC DR InterPro; IPR001881; EGF_Ca.
 CC DR Pfam; PF01414; DSL; 1.
 CC DR Pfam; PF00008; EGF; 18.
 CC DR PROSITE; PS00010; ASX_HYDROXYL; 9.
 CC DR PROSITE; PS00022; EGF 1; 16.
 CC DR PROSITE; PS01186; EGF 2; 12.
 CC DR PROSITE; PS01187; EGF_Ca; 8.
 CC DR Calcium-binding; EGF-like domain; Glycoprotein; Developmental protein;
 CC Repeat; Transmembrane; Signal.
 CC KM SIGNAL
 CC FT SIGNAL 1 28
 CC FT CHAIN 29 1242
 CC FT DOMAIN 29 1070
 CC FT TRANSMEM 1071 1095
 CC FT DOMAIN 1096 1242
 CC FT DOMAIN 168 220
 CC FT DOMAIN 221 264
 CC FT DOMAIN 264 295
 CC FT DOMAIN 297 335
 CC FT DOMAIN 337 373
 CC FT DOMAIN 373 411
 CC FT DOMAIN 411 449
 CC FT DOMAIN 449 486
 CC FT DOMAIN 486 524
 CC FT DOMAIN 524 562
 CC FT DOMAIN 562 600
 CC FT DOMAIN 600 632
 CC FT DOMAIN 632 668

FT DOMAIN 670 706 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 708 744 EGF-LIKE 13.
 FT DOMAIN 747 783 EGF-LIKE 14.
 FT DOMAIN 785 821 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 823 859 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 861 897 EGF-LIKE 17.
 FT DOMAIN 899 935 BY SIMILARITY.
 FT DISULFID 937 959 BY SIMILARITY.
 FT DISULFID 961 983 BY SIMILARITY.
 FT DISULFID 985 1007 BY SIMILARITY.
 FT DISULFID 1009 1031 BY SIMILARITY.
 FT DISULFID 1033 1055 BY SIMILARITY.
 FT DISULFID 1057 1079 BY SIMILARITY.
 FT DISULFID 1081 1103 BY SIMILARITY.
 FT DISULFID 1105 1127 BY SIMILARITY.
 FT DISULFID 1129 1151 BY SIMILARITY.
 FT DISULFID 1153 1175 BY SIMILARITY.
 FT DISULFID 1177 1199 BY SIMILARITY.
 FT DISULFID 1201 1223 BY SIMILARITY.
 FT DISULFID 1225 1247 BY SIMILARITY.
 FT DISULFID 1249 1271 BY SIMILARITY.
 FT DISULFID 1273 1295 BY SIMILARITY.
 FT DISULFID 1297 1319 BY SIMILARITY.
 FT DISULFID 1321 1343 BY SIMILARITY.
 FT DISULFID 1345 1367 BY SIMILARITY.
 FT DISULFID 1369 1391 BY SIMILARITY.
 FT DISULFID 1393 1415 BY SIMILARITY.
 FT DISULFID 1417 1439 BY SIMILARITY.
 FT DISULFID 1441 1463 BY SIMILARITY.
 FT DISULFID 1465 1487 BY SIMILARITY.
 FT DISULFID 1489 1511 BY SIMILARITY.
 FT DISULFID 1513 1535 BY SIMILARITY.
 FT DISULFID 1537 1559 BY SIMILARITY.
 FT DISULFID 1561 1583 BY SIMILARITY.
 FT DISULFID 1585 1607 BY SIMILARITY.
 FT DISULFID 1609 1631 BY SIMILARITY.
 FT DISULFID 1633 1655 BY SIMILARITY.
 FT DISULFID 1657 1679 BY SIMILARITY.
 FT DISULFID 1681 1703 BY SIMILARITY.
 FT DISULFID 1705 1727 BY SIMILARITY.
 FT DISULFID 1729 1751 BY SIMILARITY.
 FT DISULFID 1753 1775 BY SIMILARITY.
 FT DISULFID 1777 1799 BY SIMILARITY.
 FT DISULFID 1801 1823 BY SIMILARITY.
 FT DISULFID 1825 1847 BY SIMILARITY.
 FT DISULFID 1849 1871 BY SIMILARITY.
 FT DISULFID 1873 1895 BY SIMILARITY.
 FT DISULFID 1897 1919 BY SIMILARITY.
 FT DISULFID 1921 1943 BY SIMILARITY.
 FT DISULFID 1945 1967 BY SIMILARITY.
 FT DISULFID 1969 1991 BY SIMILARITY.
 FT DISULFID 1993 2015 BY SIMILARITY.
 FT DISULFID 2017 2039 BY SIMILARITY.
 FT DISULFID 2041 2063 BY SIMILARITY.
 FT DISULFID 2065 2087 BY SIMILARITY.
 FT DISULFID 2089 2111 BY SIMILARITY.
 FT DISULFID 2113 2135 BY SIMILARITY.
 FT DISULFID 2137 2159 BY SIMILARITY.
 FT DISULFID 2161 2183 BY SIMILARITY.
 FT DISULFID 2185 2207 BY SIMILARITY.
 FT DISULFID 2209 2231 BY SIMILARITY.
 FT DISULFID 2233 2255 BY SIMILARITY.
 FT DISULFID 2257 2279 BY SIMILARITY.
 FT DISULFID 2281 2303 BY SIMILARITY.
 FT DISULFID 2305 2327 BY SIMILARITY.
 FT DISULFID 2329 2351 BY SIMILARITY.
 FT DISULFID 2353 2375 BY SIMILARITY.
 FT DISULFID 2377 2399 BY SIMILARITY.
 FT DISULFID 2401 2423 BY SIMILARITY.
 FT DISULFID 2425 2447 BY SIMILARITY.
 FT DISULFID 2449 2471 BY SIMILARITY.
 FT DISULFID 2473 2495 BY SIMILARITY.
 FT DISULFID 2497 2519 BY SIMILARITY.
 FT DISULFID 2521 2543 BY SIMILARITY.
 FT DISULFID 2545 2567 BY SIMILARITY.
 FT DISULFID 2569 2591 BY SIMILARITY.
 FT DISULFID 2593 2615 BY SIMILARITY.
 FT DISULFID 2617 2639 BY SIMILARITY.
 FT DISULFID 2641 2663 BY SIMILARITY.
 FT DISULFID 2665 2687 BY SIMILARITY.
 FT DISULFID 2689 2711 BY SIMILARITY.
 FT DISULFID 2713 2735 BY SIMILARITY.
 FT DISULFID 2737 2759 BY SIMILARITY.
 FT DISULFID 2761 2783 BY SIMILARITY.
 FT DISULFID 2785 2807 BY SIMILARITY.
 FT DISULFID 2809 2831 BY SIMILARITY.
 FT DISULFID 2833 2855 BY SIMILARITY.
 FT DISULFID 2857 2879 BY SIMILARITY.
 FT DISULFID 2881 2903 BY SIMILARITY.
 FT DISULFID 2905 2927 BY SIMILARITY.
 FT DISULFID 2929 2951 BY SIMILARITY.
 FT DISULFID 2953 2975 BY SIMILARITY.
 FT DISULFID 2977 2999 BY SIMILARITY.
 FT DISULFID 3001 3023 BY SIMILARITY.
 FT DISULFID 3025 3047 BY SIMILARITY.
 FT DISULFID 3049 3071 BY SIMILARITY.
 FT DISULFID 3073 3095 BY SIMILARITY.
 FT DISULFID 3097 3119 BY SIMILARITY.
 FT DISULFID 3121 3143 BY SIMILARITY.
 FT DISULFID 3145 3167 BY SIMILARITY.
 FT DISULFID 3169 3191 BY SIMILARITY.
 FT DISULFID 3193 3215 BY SIMILARITY.
 FT DISULFID 3217 3239 BY SIMILARITY.
 FT DISULFID 3241 3263 BY SIMILARITY.
 FT DISULFID 3265 3287 BY SIMILARITY.
 FT DISULFID 3289 3311 BY SIMILARITY.
 FT DISULFID 3313 3335 BY SIMILARITY.
 FT DISULFID 3337 3359 BY SIMILARITY.
 FT DISULFID 3361 3383 BY SIMILARITY.
 FT DISULFID 3385 3407 BY SIMILARITY.
 FT DISULFID 3409 3431 BY SIMILARITY.
 FT DISULFID 3433 3455 BY SIMILARITY.
 FT DISULFID 3457 3479 BY SIMILARITY.
 FT DISULFID 3481 3503 BY SIMILARITY.
 FT DISULFID 3505 3527 BY SIMILARITY.
 FT DISULFID 3529 3551 BY SIMILARITY.
 FT DISULFID 3553 3575 BY SIMILARITY.
 FT DISULFID 3577 3599 BY SIMILARITY.
 FT DISULFID 3601 3623 BY SIMILARITY.
 FT DISULFID 3625 3647 BY SIMILARITY.
 FT DISULFID 3649 3671 BY SIMILARITY.
 FT DISULFID 3673 3695 BY SIMILARITY.
 FT DISULFID 3697 3719 BY SIMILARITY.
 FT DISULFID 3721 3743 BY SIMILARITY.
 FT DISULFID 3745 3767 BY SIMILARITY.
 FT DISULFID 3769 3791 BY SIMILARITY.
 FT DISULFID 3793 3815 BY SIMILARITY.
 FT DISULFID 3817 3839 BY SIMILARITY.
 FT DISULFID 3841 3863 BY SIMILARITY.
 FT DISULFID 3865 3887 BY SIMILARITY.
 FT DISULFID 3889 3911 BY SIMILARITY.
 FT DISULFID 3913 3935 BY SIMILARITY.
 FT DISULFID 3937 3959 BY SIMILARITY.
 FT DISULFID 3961 3983 BY SIMILARITY.
 FT DISULFID 3985 4007 BY SIMILARITY.
 FT DISULFID 4009 4031 BY SIMILARITY.
 FT DISULFID 4033 4055 BY SIMILARITY.
 FT DISULFID 4057 4079 BY SIMILARITY.
 FT DISULFID 4081 4103 BY SIMILARITY.
 FT DISULFID 4105 4127 BY SIMILARITY.
 FT DISULFID 4129 4151 BY SIMILARITY.
 FT DISULFID 4153 4175 BY SIMILARITY.
 FT DISULFID 4177 4199 BY SIMILARITY.
 FT DISULFID 4201 4223 BY SIMILARITY.
 FT DISULFID 4225 4247 BY SIMILARITY.
 FT DISULFID 4249 4271 BY SIMILARITY.
 FT DISULFID 4273 4295 BY SIMILARITY.
 FT DISULFID 4297 4319 BY SIMILARITY.
 FT DISULFID 4321 4343 BY SIMILARITY.
 FT DISULFID 4345 4367 BY SIMILARITY.
 FT DISULFID 4369 4391 BY SIMILARITY.
 FT DISULFID 4393 4415 BY SIMILARITY.
 FT DISULFID 4417 4439 BY SIMILARITY.
 FT DISULFID 4441 4463 BY SIMILARITY.
 FT DISULFID 4465 4487 BY SIMILARITY.
 FT DISULFID 4489 4511 BY SIMILARITY.
 FT DISULFID 4513 4535 BY SIMILARITY.
 FT DISULFID 4537 4559 BY SIMILARITY.
 FT DISULFID 4561 4583 BY SIMILARITY.
 FT DISULFID 4585 4607 BY SIMILARITY.
 FT DISULFID 4609 4631 BY SIMILARITY.
 FT DISULFID 4633 4655 BY SIMILARITY.
 FT DISULFID 4657 4679 BY SIMILARITY.
 FT DISULFID 4681 4703 BY SIMILARITY.
 FT DISULFID 4705 4727 BY SIMILARITY.
 FT DISULFID 4729 4751 BY SIMILARITY.
 FT DISULFID 4753 4775 BY SIMILARITY.
 FT DISULFID 4777 4799 BY SIMILARITY.
 FT DISULFID 4801 4823 BY SIMILARITY.
 FT DISULFID 4825 4847 BY SIMILARITY.
 FT DISULFID 4849 4871 BY SIMILARITY.
 FT DISULFID 4873 4895 BY SIMILARITY.
 FT DISULFID 4897 4919 BY SIMILARITY.
 FT DISULFID 4921 4943 BY SIMILARITY.
 FT DISULFID 4945 4967 BY SIMILARITY.
 FT DISULFID 4969 4991 BY SIMILARITY.
 FT DISULFID 4993 5015 BY SIMILARITY.
 FT DISULFID 5017 5039 BY SIMILARITY.
 FT DISULFID 5041 5063 BY SIMILARITY.
 FT DISULFID 5065 5087 BY SIMILARITY.
 FT DISULFID 5089 5111 BY SIMILARITY.
 FT DISULFID 5113 5135 BY SIMILARITY.
 FT DISULFID 5137 5159 BY SIMILARITY.
 FT DISULFID 5161 5183 BY SIMILARITY.
 FT DISULFID 5185 5207 BY SIMILARITY.
 FT DISULFID 5209 5231 BY SIMILARITY.
 FT DISULFID 5233 5255 BY SIMILARITY.
 FT DISULFID 5257 5279 BY SIMILARITY.
 FT DISULFID 5281 5303 BY SIMILARITY.
 FT DISULFID 5305 5327 BY SIMILARITY.
 FT DISULFID 5329 5351 BY SIMILARITY.
 FT DISULFID 5353 5375 BY SIMILARITY.
 FT DISULFID 5377 5399 BY SIMILARITY.
 FT DISULFID 5401 5423 BY SIMILARITY.
 FT DISULFID 5425 5447 BY SIMILARITY.
 FT DISULFID 5449 5471 BY SIMILARITY.
 FT DISULFID 5473 5495 BY SIMILARITY.
 FT DISULFID 5497 5519 BY SIMILARITY.
 FT DISULFID 5521 5543 BY SIMILARITY.
 FT DISULFID 5545 5567 BY SIMILARITY.
 FT DISULFID 5569 5591 BY SIMILARITY.
 FT DISULFID 5593 5615 BY SIMILARITY.
 FT DISULFID 5617 5639 BY SIMILARITY.
 FT DISULFID 5641 5663 BY SIMILARITY.
 FT DISULFID 5665 5687 BY SIMILARITY.
 FT DISULFID 5689 5711 BY SIMILARITY.
 FT DISULFID 5713 5735 BY SIMILARITY.
 FT DISULFID 5737 5759 BY SIMILARITY.
 FT DISULFID 5761 5783 BY SIMILARITY.
 FT DISULFID 5785 5807 BY SIMILARITY.
 FT DISULFID 5809 5831 BY SIMILARITY.
 FT DISULFID 5833 5855 BY SIMILARITY.
 FT DISULFID 5857 5879 BY SIMILARITY.
 FT DISULFID 5881 5903 BY SIMILARITY.
 FT DISULFID 5905 5927 BY SIMILARITY.
 FT DISULFID 5929 5951 BY SIMILARITY.
 FT DISULFID 5953 5975 BY SIMILARITY.
 FT DISULFID 5977 5999 BY SIMILARITY.
 FT DISULFID 6001 6023 BY SIMILARITY.
 FT DISULFID 6025 6047 BY SIMILARITY.
 FT DISULFID 6049 6071 BY SIMILARITY.
 FT DISULFID 6073 6095 BY SIMILARITY.
 FT DISULFID 6097 6119 BY SIMILARITY.
 FT DISULFID 6121 6143 BY SIMILARITY.
 FT DISULFID 6145 6167 BY SIMILARITY.
 FT DISULFID 6169 6191 BY SIMILARITY.
 FT DISULFID 6193 6215 BY SIMILARITY.
 FT DISULFID 6217 6239 BY SIMILARITY.
 FT DISULFID 6241 6263 BY SIMILARITY.
 FT DISULFID 6265 6287 BY SIMILARITY.
 FT DISULFID 6289 6311 BY SIMILARITY.
 FT DISULFID 6313 6335 BY SIMILARITY.
 FT DISULFID 6337 6359 BY SIMILARITY.
 FT DISULFID 6361 6383 BY SIMILARITY.
 FT DISULFID 6385 6407 BY SIMILARITY.
 FT DISULFID 6409 6431 BY SIMILARITY.
 FT DISULFID 6433 6455 BY SIMILARITY.
 FT DISULFID 6457 6479 BY SIMILARITY.
 FT DISULFID 6481 6503 BY SIMILARITY.
 FT DISULFID 6505 6527 BY SIMILARITY.
 FT DISULFID 6529 6551 BY SIMILARITY.
 FT DISULFID 6553 6575 BY SIMILARITY.
 FT DISULFID 6577 6599 BY SIMILARITY.
 FT DISULFID 6601 6623 BY SIMILARITY.
 FT DISULFID 6625 6647 BY SIMILARITY.
 FT DISULFID 6649 6671 BY SIMILARITY.
 FT DISULFID 6673 6695 BY SIMILARITY.
 FT DISULFID 6697 6719 BY SIMILARITY.
 FT DISULFID 6721 6743 BY SIMILARITY.
 FT DISULFID 6745 6767 BY SIMILARITY.
 FT DISULFID 6769 6791 BY SIMILARITY.
 FT DISULFID 6793 6815 BY SIMILARITY.
 FT DISULFID 6817 6839 BY SIMILARITY.
 FT DISULFID 6841 6863 BY SIMILARITY.
 FT DISULFID 6865 6887 BY SIMILARITY.
 FT DISULFID 6889 6911 BY SIMILARITY.
 FT DISULFID 6913 6935 BY SIMILARITY.
 FT DISULFID 6937 6959 BY SIMILARITY.
 FT DISULFID 6961 6983 BY SIMILARITY.
 FT DISULFID 6985 7007 BY SIMILARITY.
 FT DISULFID 7009 7031 BY SIMILARITY.
 FT DISULFID 7033 7055 BY SIMILARITY.
 FT DISULFID 7057 7079 BY SIMILARITY.
 FT DISULFID 7081 7103 BY SIMILARITY.
 FT DISULFID 7105 7127 BY SIMILARITY.
 FT DISULFID 7129 7151 BY SIMILARITY.
 FT DISULFID 7153 7175 BY SIMILARITY.
 FT DISULFID 7177 7199 BY SIMILARITY.
 FT DISULFID 7201 7223 BY SIMILARITY.
 FT DISULFID 7225 7247 BY SIMILARITY.
 FT DISULFID 7249 7271 BY SIMILARITY.
 FT DISULFID 7273 7295 BY SIMILARITY.
 FT DISULFID 7297 7319 BY SIMILARITY.
 FT DISULFID 7321 7343 BY SIMILARITY.
 FT DISULFID 7345 7367 BY SIMILARITY.
 FT DISULFID 7369 7391 BY SIMILARITY.
 FT DISULFID 7393 7415 BY SIMILARITY.
 FT DISULFID 7417 7439 BY SIMILARITY.
 FT DISULFID 7441 7463 BY SIMILARITY.
 FT DISULFID 7465 7487 BY SIMILARITY.
 FT DISULFID 7489 7511 BY SIMILARITY.
 FT DISULFID 7513 7535 BY SIMILARITY.
 FT DISULFID 7537 7559 BY SIMILARITY.
 FT DISULFID 7561 7583 BY SIMILARITY.
 FT DISULFID 7585 7607 BY SIMILARITY.
 FT DISULFID 7609 7631 BY SIMILARITY.
 FT DISULFID 7633 7655 BY SIMILARITY.
 FT DISULFID 7657 7679 BY SIMILARITY.
 FT DISULFID 7681 7703 BY SIMILARITY.
 FT DISULFID 7705 7727 BY SIMILARITY.
 FT DISULFID 7729 7751 BY SIMILARITY.
 FT DISULFID 7753 7775 BY SIMILARITY.
 FT DISULFID 7777 7799 BY SIMILARITY.
 FT DISULFID 7801 7823 BY SIMILARITY.
 FT DISULFID 7825 7847 BY SIMILARITY.
 FT DISULFID 7849 7871 BY SIMILARITY.
 FT DISULFID 7873 7895 BY SIMILARITY.
 FT DISULFID 7897 7919 BY SIMILARITY.
 FT DISULFID 7921 7943 BY SIMILARITY.
 FT DISULFID 7945 7967 BY SIMILARITY.
 FT DISULFID 7969 7991 BY SIMILARITY.
 FT DISULFID 7993 8015 BY SIMILARITY.
 FT DISULFID 8017 8039 BY SIMILARITY.
 FT DISULFID 8041 8063 BY SIMILARITY.
 FT DISULFID 8065 8087 BY SIMILARITY.
 FT DISULFID 8089 8111 BY SIMILARITY.
 FT DISULFID 8113 8135 BY SIMILARITY.
 FT DISULFID 8137 8159 BY SIMILARITY.
 FT DISULFID 8161 8183 BY SIMILARITY.
 FT DISULFID 8185 8207 BY SIMILARITY.
 FT DISULFID 8209 8231 BY SIMILARITY.
 FT DISULFID 8233 8255 BY SIMILARITY.
 FT DISULFID 8257 8279 BY SIMILARITY.
 FT DISULFID 8281 8303 BY SIMILARITY.
 FT DISULFID 8305 8327 BY SIMILARITY.
 FT DISULFID 8329 8351 BY SIMILARITY.
 FT DISULFID 8353 8375 BY SIMILARITY.
 FT DISULFID 8377 8399 BY SIMILARITY.
 FT DISULFID 8401 8423 BY SIMILARITY.
 FT DISULFID 8425 8447 BY SIMILARITY.
 FT DISULFID 8449 8471 BY SIMILARITY.
 FT DISULFID 8473 8495 BY SIMILARITY.
 FT DISULFID 8497 8519 BY SIMILARITY.
 FT DISULFID 8521 8543 BY SIMILARITY.
 FT DISULFID 8545 8567 BY SIMILARITY.
 FT DISULFID 8569 8591 BY SIMILARITY.
 FT DISULFID 8593 8615 BY SIMILARITY.
 FT DISULFID 8617 8639 BY SIMILARITY.
 FT DISULFID 8641 8663 BY SIMILARITY.
 FT DISULFID 8665 8687 BY SIMILARITY.
 FT DISULFID 8689 8711 BY SIMILARITY.
 FT DISULFID 8713 8735 BY SIMILARITY.
 FT DISULFID 8737 8759 BY SIMILARITY.
 FT DISULFID 8761 8783 BY SIMILARITY.
 FT DISULFID 8785 8807 BY SIMILARITY.
 FT DISULFID 8809 8831 BY SIMILARITY.
 FT DISULFID 8833 8855 BY SIMILARITY.
 FT DISULFID 8857 8879 BY SIMILARITY.
 FT DISULFID 8881 8903 BY SIMILARITY.
 FT DISULFID 8905 8927 BY SIMILARITY.
 FT DISULFID 8929 8951 BY SIMILARITY.
 FT DISULFID 8953 8975 BY SIMILARITY.
 FT DISULFID 8977 8999 BY SIMILARITY.
 FT DISULFID 9001 9023 BY SIMILARITY.
 FT DISULFID 9025 9047 BY SIMILARITY.
 FT DISULFID 9049 9071 BY SIMILARITY.
 FT DISULFID 9073 9095 BY SIMILARITY.
 FT DISULFID 9097 9119 BY SIMILARITY.
 FT DISULFID 9121 9143 BY SIMILARITY.
 FT DISULFID 9145 9167 BY SIMILARITY.
 FT DISULFID 9169 9191 BY SIMILARITY.
 FT DISULFID 9193 9215 BY SIMILARITY.
 FT DISULFID 9217 9239 BY SIMILARITY.
 FT DISULFID 9241 9263 BY SIMILARITY.
 FT DISULFID 9265 9287 BY SIMILARITY.
 FT DISULFID 9289 9311 BY SIMILARITY.
 FT DISULFID 9313 9335 BY SIMILARITY.
 FT DISULFID 9337 9359 BY SIMILARITY.
 FT DISULFID 9361 9383 BY SIMILARITY.
 FT DISULFID 9385 9407 BY SIMILARITY.
 FT DISULFID 9409 9431 BY SIMILARITY.
 FT DISULFID 9433 9455 BY SIMILARITY.
 FT DISULFID 9457 9479 BY SIMILARITY.
 FT DISULFID 9481 9503 BY SIMILARITY.
 FT DISULFID 9505 9527 BY SIMILARITY.
 FT DISULFID 9529 9551 BY SIMILARITY.
 FT DISULFID 9553 9575 BY SIMILARITY.
 FT DISULFID 9577 9599 BY SIMILARITY.
 FT DISULFID 9601 9623 BY SIMILARITY.
 FT DISULFID 9625 9647 BY SIMILARITY.
 FT DISULFID 9649 9671 BY SIMILARITY.
 FT DISULFID 9673 9695 BY SIMILARITY.
 FT DISULFID 9697 9719 BY SIMILARITY.
 FT DISULFID 9721 9743 BY SIMILARITY.
 FT DISULFID 9745 9767 BY SIMILARITY.
 FT DISULFID 9769 9791 BY SIMILARITY.
 FT DISULFID 9793 9815 BY SIMILARITY.
 FT DISULFID 9817 9839 BY SIMILARITY.
 FT DISULFID 9841 9863 BY SIMILARITY.
 FT DISULFID 9865 9887 BY SIMILARITY.
 FT DISULFID 9889 9911 BY SIMILARITY.
 FT DISULFID 9913 9935 BY SIMILARITY.
 FT DISULFID 9937 9959 BY SIMILARITY.
 FT DISULFID 9961 9983 BY SIMILARITY.
 FT DISULFID 9985 10007 BY SIMILARITY.

Query Match 7.2%; Score 145; DB 1; Length 1242;
 Best Local Similarity 21.1%; Pred. No. 0.017;
 Matches 82; Conservative 45; Mismatches 142; Indels 120; Gaps 22;

QY 32 OCGGND-WNGPTC-----CEGSGTCK-VSNPYSGCLAPESNGNSSECSKLYG 78
 Db 655 QCIADWEGSPNCEFTNDDCTNPDRRGVCRDLVNDYCEC---ENWKGKCTHSRES 710
 QY 79 QG-----GG-----KQWNGPTC-----CEGSGTCKVSNPYYS 105
 Db 711 QGDEDTQNNGGTCSDEBGSFCKLCSPGWEGATCNIAKNSCLPNCENGATCVTGDGFT 770

```

Qy 106 QCLAPESNGNKTSESANHTTTTAPAKEITTTAKANSNSNSGKYIIVGGASGNGVTTR 165
Db 771 -CVCKE-----GMEGPTCSQNSNDNCNPQ?CYN-----SGTCVDGD 804
Qy 166 YMDCKKASCWPG---KANVSSPVKSCNKDGVTLSDSN-VQSGCNGGNS-ymcnd--NQ 218
Db 805 NMYRCECASGFAGPDCRININECOSSPCAFSGTCVDEINGYRCLCPGRTGPRCQEVYGR 864
Qy 219 PMAVNDNLAYGFPAALISGGESRWC--CSCEELFTSTV-AGKMWIQVNTNGDLGS 275
Db 865 PCVIGGRI-----AVDGAHMAEDCNTCYCHKGIYCTKLP?GPKACRMILSGGRD--C 915
Qy 276 STGAHFDLQMPGGGVGIF-NGCSKQWGA?NDGWSRYGJISSASDCSSLPSALQAGCKMR 334
Db 916 PTG---QLCVFVRDEQCFVKPCSSQ---GECMSAHRPAVRTHCQPPDS----- 956
Qy 335 FNMFRKADNPSMTYKEVTCPEKITAKTGC 363
Db 957 -----HCANVTFTFNKDTMQGVTVEQVC 980

```

Search completed: June 18, 2003, 15:32:03
 Job time : 11.6 secs

7

THIS PAGE BLANK (USPTO)